AD-A102 567 CALIFORNIA UNIV BERKELEY CENTER FOR PURE AND APPLIE--ETC F/G 12/1
HOW TO SOLVE (K - LAMBDA M)Z = 0 FOR LARGE K AND M,(U)
JUN 81 B N PARLETT N00014-76-C-0013
UNCLASSIFIED PAM-39 NL

END
DATE FILMED
9-81
DTIC

LEVEL II

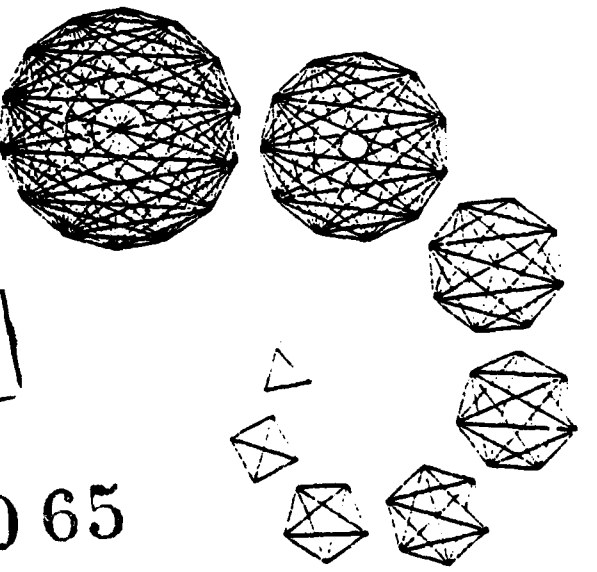
CENTER FOR PURE AND APPLIED MATHEMATICS
UNIVERSITY OF CALIFORNIA, BERKELEY

PAM-39

HOW TO SOLVE $(K - \lambda M)z = 0$ FOR LARGE K AND M

B. N. PARLETT

26 June 81
(per letter)



Legal Notice

14 [illegible]-39

11 [illegible]

12 [illegible]

6 lambda

HOW TO SOLVE $(K - \lambda M)z = 0$ FOR LARGE K AND M

B.N. Parlett(*)

10

Abstract

The most common solution technique is subspace iteration, a combination of inverse iteration and the Rayleigh-Ritz procedure. Some difficulties of this method are mentioned as well as ways to avoid them by means of spectral transformations. This permits use of the Lanczos algorithm and yields significant reductions in cost.

This paper evolved from an invited talk at the 2nd International Congress on Numerical Methods in Engineering (GAMNI 2) at the Ecole Centrale de Paris in December 1980.

(*) Mathematics Department and Computer Sciences Division, University of California, Berkeley. The author gratefully acknowledges support from Office of Naval Research Contract N00014-76-C-0013.

13

404935

1. INTRODUCTION

In order to avoid excessive generality we assume that K and M are real, symmetric $n \times n$ matrices. The goal is to compute all the eigenvalues lying in a given interval $[\tau_1, \tau_2]$ together with the corresponding eigenvectors.

For $n > 2$ the eigenvalues will be real provided that $\kappa K + \mu M$ is positive definite for some choice of κ and μ. For simplicity assume that either K or M is positive definite.

Example. If $K = M = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ then every number (real or complex) is an eigenvalue belonging to $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

The pair (K, M) is often called a pencil. Its eigenvalues and eigenvectors are denoted by

$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n ,$$

$$z_1, z_2, \ldots, z_n .$$

Techniques for solving the problem when $n < 100$ are described in [Parlett, 1980] and will not be mentioned here.

The most important factor in selecting a particular algorithm is the relative cost of solving $(K - \sigma M)u = v$ for u to forming the product $u = (K - \sigma M)v$. Observe that this is simply the divide: multiply ratio for the given matrix pencil. Linear equations can be solved iteratively

(using conjugate gradients or Lanczos) when triangular factorization is not possible. The remaining sections show how this divide: multiply ratio affects the solution technique.

With apologies to structural engineers we use capital letters for matrices, small roman letters for column vectors, and small greek letters for numbers.

2. SUBSPACE ITERATION

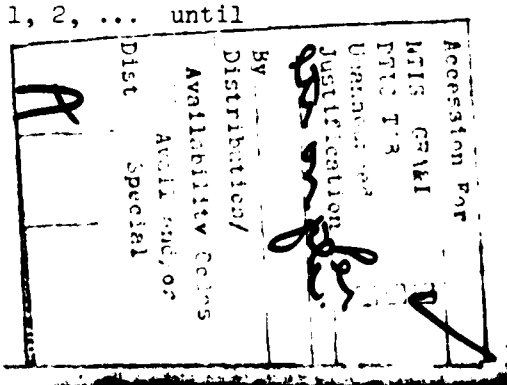
See [Bathé and Wilson, 1976] or [Parlett, 1980] for more details. It is assumed that M is positive definite.

Before the start, a shift σ is chosen in $[\tau_1, \tau_2]$ and the triangular factorization $K - \sigma M = LDL^*$ is computed. Note that for each $i = 1, \ldots, n$,

$$(K - \sigma M)^{-1} M z_i = z_i \nu_i \ , \quad \nu_i = \frac{1}{(\lambda_i - \sigma)} \ . \tag{2-1}$$

So the power method with the operator $(K - \sigma M)^{-1} M$ would converge to the eigenvector z_i belonging to the λ_i closest to σ. The usual normalization is $z_i^* M z_j = \delta_{ij}$ (Kronecker). Of course, $K - \sigma M$ is <u>not</u> to be inverted explicitly.

The number of eigenvalues in $[\tau_1, \tau_2]$ is unknown but subspace iteration must begin with the difficult choice of the dimension p and the somewhat easier choice of starting vectors $(x_1, x_2, \ldots, x_p) \equiv X_0$ with $x_0^* M X_0 = I_p$. The usual implementation is: for $k = 1, 2, \ldots$ until convergence repeat



1. Solve $(K - \sigma M)Y_k = MX_{k-1}$ for Y_k .
2. Compute the reduced matrices (or projections) $K_k = Y_k{}^*KY_k$, $M_k = Y_k{}^*MY_k$.
3. Solve the small problem $(K_k - \theta_i M_k)g_i = 0$ for eigenvalues θ_i and normalized eigenvectors $g_1, \ldots, g_p$. Set $G_k \equiv (g_1, \ldots, g_p)$.
4. Set $X_k = Y_k G_k$.
5. Test for convergence.

To test for convergence one should use

> Theorem. For any $x \neq 0$ and any scalar θ there exists an eigenvalue of (K, M) such that
>
> $$|\lambda - \theta| \leq \|(K - \theta M)x\|_{M^{-1}} / \|Mx\|_{M^{-1}} .$$

Here $\|u\|_H = \sqrt{u^*Hu}$. Thus $\|Mx\|_{M^{-1}} = \sqrt{x^*Mx}$ which is easy to compute but the cost of the numerator discourages the use of this theorem. Often users wait until the values of θ "settle down" to the desired accuracy but such a criterion is not completely reliable.

A serious difficulty with subspace iteration occurs when (K, M) has eigenvalues outside <u>both</u> ends of $[\tau_1, \tau_2]$. To illustrate what happens consider an extreme case: suppose that at some step k one of the Ritz vectors (the columns of X_k) is x and satisfies $x = (z_1 + z_{30})/\sqrt{2}$ and

$$\lambda_1 < \tau_1 < \tau_2 < \lambda_{30} \; ; \quad \lambda_1 + \lambda_{30} = \tau_1 + \tau_2 .$$

It turns out that the corresponding Ritz value is

$$\theta = \frac{z_1^{*}Mz_1 \cdot \lambda_1 + z_{30}^{*}Mz_{30} \cdot \lambda_{30}}{z_1^{*}Mz_1 + z_2^{*}Mz_2} = \frac{\lambda_1+\lambda_{30}}{2} = \frac{\tau_1+\tau_2}{2} .$$

Although θ is in the middle of $[\tau_1, \tau_2]$ it does not signal that there is an eigenvalue λ nearby. R.L. Taylor calls these "ghost" values and in [Scott, 1981] it is shown how their presence can prevent the discovery of good approximate eigenvectors. One remedy is to compute all eigenvalues and eigenvectors less than τ_1 and then keep X_k orthogonal to these eigenvectors. That is expensive.

3. TRANSFORMING THE SPECTRUM

Steps 2, 3, and 4 of subspace iteration carry out the Raleigh-Ritz procedure which delivers the best approximations to eigenvalues of (K, M) from the space spanned by Y_k. To remove the troubles of Section 2 apply the Rayleigh-Ritz procedure to $((K - \sigma M)^{-1}M, I)$ because the wanted eigenvalues ν_i, see (2-1), are the extreme ones. However $(K - \sigma M)^{-1}M$ is not symmetric. There are two ways out of this difficulty.

I. <u>The Spectral Transformation</u> (see [Ericsson and Ruhe, 1980] for details).

Factor $M = LL^{*}$ and use $L^{*}(K - \sigma M)^{-1}L$ instead of $(K - \sigma M)^{-1}M$.
In other words change $(K - \lambda M)z = 0$ into

$$\frac{1}{\lambda-\sigma}(L^{*}z) = L^{*}(K - \sigma M)^{-1}L(L^{*}z) . \qquad (3\text{-}1)$$

The powerful Lanczos algorithm can be used with the new operator. Another advantage is that the error bound of Section 2 becomes

$$|\theta - \nu| \leq \frac{\|L^*(K-\sigma M)^{-1}Lx - x\theta\|}{x} \qquad (3\text{-}2)$$

and the residual vectors for X_{k-1} are available after part 1 of step k.

The price paid for these advantages is that M must be factored although it need not be invertible. It is also preferable to factor $K - \sigma M$ if this is feasible.

II. The inertial inner product.

Recall that we want to use the operator, or matrix, $(K - \sigma M)^{-1}M$ because its eigenvalue distribution is advantageous for both subspace iteration and for the Lanczos algorithm. However $(K - \sigma M)^{-1}M$ is not symmetric and it is often said that Lanczos requires a symmetric matrix. That is not quite correct.

Let us assume for the moment that M is positive definite. We will consider singular M at the end of this section. The key observation is that $(K - \sigma M)^{-1}M$ is self-adjoint with respect to the "inertial" inner product

$$(u, v)_M \equiv v^*Mu \ .$$

Here is the proof.

$$\begin{aligned} ((K - \sigma M)^{-1}Mu, v)_M &= v^*M\cdot(K - \sigma M)^{-1}Mu \ , \\ &= v^*M(K - \sigma M)^{-1}\cdot Mu \ , \\ &= (u, (K - \sigma M)^{-1}Mv)_M \ . \quad \square \end{aligned}$$

As usual $\|u\|_M \equiv \sqrt{(u,u)}_M$. Here is the algorithm.

Lanczos Algorithm [C must be self-adjoint with respect to $(\cdot, \cdot)_M$] .

Pick $r_1 \neq 0$ and compute $\beta_1 = \|r_1\|_M$. Set $q_0 = 0$.

For $j = 1, 2, \ldots$ until convergence repeat

1. $q_j = r_j/\beta_j$
2. $u_j = Cq_j - q_{j-1}\beta_j$
3. $\alpha_j = (q_j, u_j)_M$
4. $r_{j+1} = u_j - q_j\alpha_j$
5. $\beta_{j+1} = \|r_{j+1}\|_M$
6. Test for convergence.

In our case $C = (K - \sigma M)^{-1}M$.

This formulation is not widely appreciated. It was given explicitly in [van Kats and van der Vorst, 1977]. An alternative, but more expensive version of the algorithm is given in [Parlett, 1980, p. 324].

The numbers α_i , β_i computed in the course of the algorithm are put into a tridiagonal matrix

$$T_j = \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \beta_2 & \alpha_2 & \beta_3 & & \\ & \beta_3 & \alpha_3 & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \alpha_j \end{bmatrix}$$

whose order j grows by one at each step.

The eigenvalues θ_i , $i = 1, \ldots, j$ of T_j are the Ritz approximations to the eigenvalues of $C = (K - M)^{-1}M$. As j increases the outer values θ_i quickly converge to the outer eigenvalues of $(K - \sigma M)^{-1}M$ and these may be converted, by $(2 - 1)$, into the required eigenvalues of (K, M) close to σ .

Let us consider now the case when M is singular but K is positive definite. It follows that (K, M) has one or more infinite eigenvalues while $C = (K - \sigma M)^{-1}M$ has one or more zero eigenvalues. Formally the bilinear expression $(\cdot, \cdot)_M$ is not a true inner product, however when restricted to the space corresponding to the finite eigenvalues of (K, M) it is an inner product. The algorithm given above automatically keeps the q_i , $i > 1$, in this subspace. It is necessary to take $r_1 = Mr_0$ to ensure that q_1 is also in the subspace. This is the only modification needed when M is positive semidefinite.

The ghost values which can afflict the Ritz approximations for the pencil (K, M) cannot affect either $(L(K - \sigma M)^{-1}L^T, I)$ or (C, I) . Consider again the example in Section 1. The Rayleigh quotient of x with respect to both pencils is $\rho = \frac{1}{2}\left(\frac{1}{\lambda_1 - \sigma} + \frac{1}{\lambda_{30} - \sigma}\right)$. For any $\sigma \in (\tau_1, \tau_2)$ we find $\rho \in \left(\frac{1}{\tau_1 - \sigma}, \frac{1}{\tau_2 - \sigma}\right)$ while the algorithm only looks at those Ritz values <u>outside</u> this interval. The fact that Ritz approximations for inverted operators prevent the occurrence of ghost vectors was pointed out in [Scott, 1981].

4. THE LANCZOS ALGORITHM

Figures 1 and 2 show a comparison of the simple Lanczos algorithm and subspace iteration on a structural problem. Comparison was stopped when subspace iteration exhausted our resources.

To be fair the simple Lanczos algorithm should be compared with the simple power method and the block Lanczos algorithm should be compared with supspace interation using the same block size. We now give a theoretical comparison to complement the the practical one.

Let A have eighenvalues $\lambda_i = i$ for $i = 0, \ldots, n + 1$. Thus A is $(n + 2)$ by $(n + 2)$. Table 1 gives the number steps m required to reduce the error in the approximation to z_{n+1} below 1% of its original value. The eigenvalue error $\lambda_{n+1} - \vartheta_m$ will then have been reduced to .01% of its original value. We assume that the initial vector makes an angle of $45°$ with the dominant eigenvector. The values for Lanczos are overestimates. The eigenvalue distribution is a difficult one but it illustrates the power of the Lanczos algorithm.

Table 1. Number of steps m to reduce error.

n	Power Method	Lanczos
10^2	463	32
10^3	4607	110
10^4	46054	380
n	n ℓn 100	$\frac{1}{2}\sqrt{n}$ ℓn 200

The superiority of even the simple Lanczos algorithm over subspace interation can be explained as follows. Suppose that subspace iteration uses a subspace of dimension p and runs for j steps. It produces Rayleigh Ritz approximation from a subspace of dimension p at every step. On the other hand, with the same computational effort

and using the same inverted operator

Lanczos produces, in exact arithmetic, the Rayleigh Ritz approximations from a subspace of dimension jp. Lanczos never forgets any vector that has been computed in the iteration, although the vector itself may have been discarded the essential information is cleverly condensed in the tridiagonal matrix T.

There is no space here to give a detailed account of the Lanczos algorithm and the reader is referred to [Parlett, 1980] for more information.

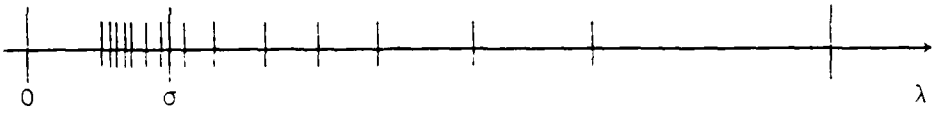
Between 1950 and 1972 the algorithm had a poor reputation, not because it was poor but because it was poorly understood. As a result of the analysis of Paige the situation was remedied and it did not take long for some good implementations to appear. One important point is that fairly cheap error bounds can be computed during the iteration. This permits the correct information to be extracted from the computed quantities and it permits the algorithm to be stopped as soon as possible.

Roundoff error has a significant impact on the Lanczos algorithm. Its effect is not to prevent convergence but only to delay it a little and also to make the extraction of eigenvalues from the Ritz values somewhat more complicated.

The importance of inverting the proper operator, implicitly of course, as discussed in Section 3, is well illustrated by the Lanczos process. There are many problems in which K and M are both positive definite, M is diagonal or nearly so, while K has a significant bandwidth (such as $\sqrt{n}$). Most text books and references suggest the following reduction to standard form

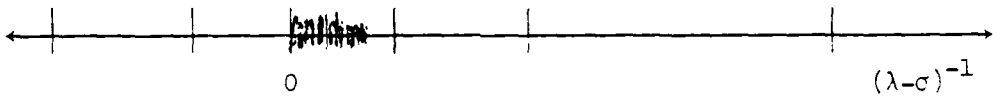
$$(K, M) \rightarrow (M^{-1/2}(K - \sigma M)M^{-1/2}, I) .$$

Usually the eigenvalues near 0 are wanted and σ may well be 0 or very small relative to the largest eigenvalue. The eigenvalue distribution typically is



This is difficult for Lanczos and disastrous for the power method with the same operator. The only attraction of this approach for the large problems is that no matrix factorization of $K - \sigma M$ is needed.

The temptation to avoid such factoring should be resisted. Even if the cost is high the reduction of $K - \sigma M$ to LDL^* is amply rewarded as suggested in Section 3. For the operator $M^{1/2}(K - \sigma M)^{-1} M^{1/2}$ the spectrum is inverted to:



This distribution is favorable for both subspace iteration and Lanczos but Lanczos takes stronger advantage it.

STORAGE CONSIDERATIONS. Lanczos requires in the fast store, (i) two or three n-vectors for the iteration, (ii) two vectors of modest length (say $5\sqrt{n}$) to hold the tridiagonal T_j , (iii) a small array to hold Ritz values. It is also desireable, but not absolutely necessary to hold in fast storage whatever is necessary to form Mq and to solve $(K - \sigma M)r = Mq$. Often this will be M and the triangular factors L and D of $K - \sigma M$. The

attractive feature of Lanczos is that each Lanczos vector can be put into secondary storage (a disk perhaps) one step after it is created. These vectors are not needed again to compute eigenvalues but they are required at the end, after convergence, to form the eigenvectors. They can also be used from time to time to improve the orthogonality among subsequent Lanczos vectors and thus to hasten convergence. This device is called selective orthogonalization and is described in [Parlett, 1980].

SHIFTS OF ORIGIN. Sometimes it is feasible to factor $K - \sigma M$ more than once, perhaps for five different values of σ. This facility helps both subspace iteration and Lanczos because the big interval $[\tau_1, \tau_2]$ can be split into five smaller ones and either algorithm can be used to find just the eigenvalues in the subinterval. The choice for the five origin shifts can be determined during the computation. Since Lanczos has available more information on the spectrum than does subspace iteration it can make somewhat better choices for the σ.

5. SOLVING $(K - \sigma M)r = Mu$.

Structural engineers are very fond of what they call (variable) band-solvers. For two dimensional problems K has a modest bandwidth and L inherits this structure. Each partial column of L is thought of as a conventional vector from the top nonzero element in the column down to the diagonal. In other words, zero elements within the band take up storage space. The gain is a beautifully simple data structure for L but for

three dimensional problems and for complicated two dimensional structures the waste of storage space is too expensive. More complicated but more effcient data structures are called for.

Considerable progress has been made in this area of sparse matrix technology and users should be aware of the enormous gains to be had from using an ordering algorithm which effectively permutes rows and columns of K and M to nearly minimize the number of nonzero elements in L. Important algorithms for this task are (i) the minimum degree algorithm, (ii) nested dissection, (iii) the Gibbs-Poole-Stockmeyer profile reducer. Usually these are all better than the standard reverse Cuthill-McKee algorithm. It is probably best to use some of the good sparse solver codes now available rather than attempt to program the algorithms oneself. The three best known codes are the Harwell library MA17 routines, SPARSPAK by Alan George and J. Liu at Univ. of Waterloo, and the Yale Sparse Matrix Package. The most recent reviews are [Duff, 1979] and [Duff and Stewart, 1979].

It is important to remember that the linear systems to be solved should be indefinite but will have most of their eigenvalues positive.

6. DAVIDSON'S METHOD

For very large $(n \geqslant 10^4)$ problems which occur in atomic and molecular orbit calculations the matrices are not very sparse but they are quite special in the sense that the eigenvector matrix is fairly close to the identity. Davidson's method [Davidson, 1975], which we have no space to describe here, is based on perturbation theory and seems to be

very effective. However, both theoretical and practical considerations [Kalamboukis, 1980] suggest that it cannot be regarded as a general purpose eigenvalue method.

REFERENCES

[1] K.-J. Bathé and E. Wilson, "Numerical Methods in Finite Element Analysis," (Prentice Hall, New Jersey).

[2] E.R. Davidson [1975], "The Iterative Calculation of a few of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real Symmetric Matrices," Jour. Comp. Phys. 17: 87-94.

[3] I.S. Duff [1979], "Some current approaches to the solution of large sparse systems of linear equations," Harwell Report CSS 65.

[4] I.S. Duff and G.W. Stewart [1979]. Sparse Matrix Proceedings 1978, SIAM Press.

[5] S.C. Eisenstat, M.H. Schultz, and A.H. Sherman [1979], "Software for Sparse Gaussian Elimination with Limited Core Storage," in Duff and Stewart, 1979.

[6] T.Z. Kalamboukis [1980], "Davidson's algorithm with and without perturbation corrections," J. Phys. A: Math. Gen. 13, 57-62.

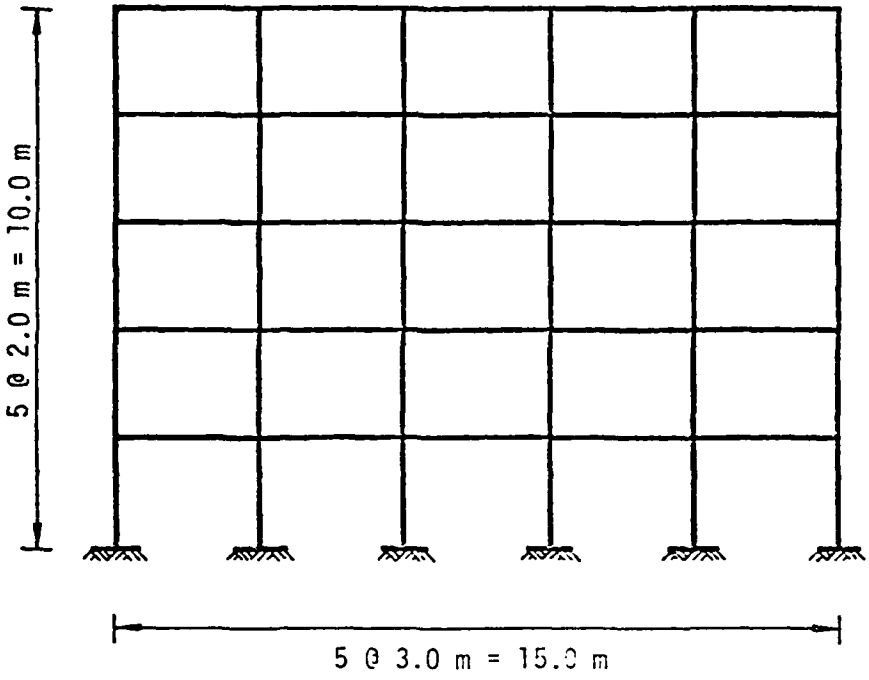
[7] B. Nour-Ohmid, B.N. Parlett, and R.L. Taylor [1981], "Lanczos Versus Subspace Iteration for Solution of Eigenvalue Problems," to appear in Int. Jour. Num. Methods in Engineering.

[8] B.N. Parlett [1980], "The Symmetric Eigenvalue Problem," Prentice Hall, New Jersey.

[9] A. Ruhe and T. Ericsson [1980], "The Spectral Transformation Lanczos Method in the Numerical Solution of Large, Sparse, Generalized, Symmetric Eigenvalue Problems," Math. Comp. 34.

[10] D.S. Scott [1981], "The Advantages of Inverted Operators in Rayleigh-Ritz Approximations," to appear.

[11] J.M. van Kats and H.A. van der Vorst [1977], "Automatic monitoring of Lanczos schemes of symmetric or skew-symmetric generalized eigenvalue problems," Tech. Report No. 7 from the Academisch Computer Centrum of Utrecht, the Natherlands.



For all beams and columns:

Young's Modulus = 1.0 KN/m^2

Mass Density = 1.0 Kg/m^3

Area = 1.0 m^2

Moment of Inertia = 1.0 m^4

No. of Beam Elements = 55

No. of Nodes = 36

Total No. of D.O.F. = 90

Figure 1a. Building Frame.

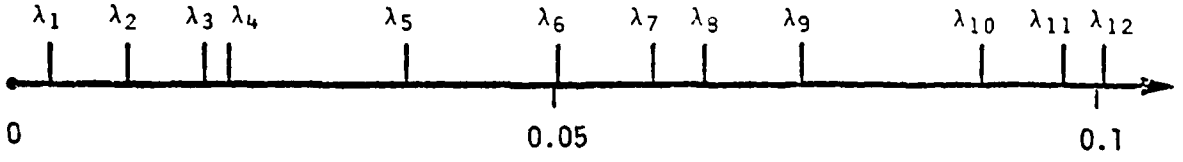


Figure 1b. Eigenvalues of the above system.

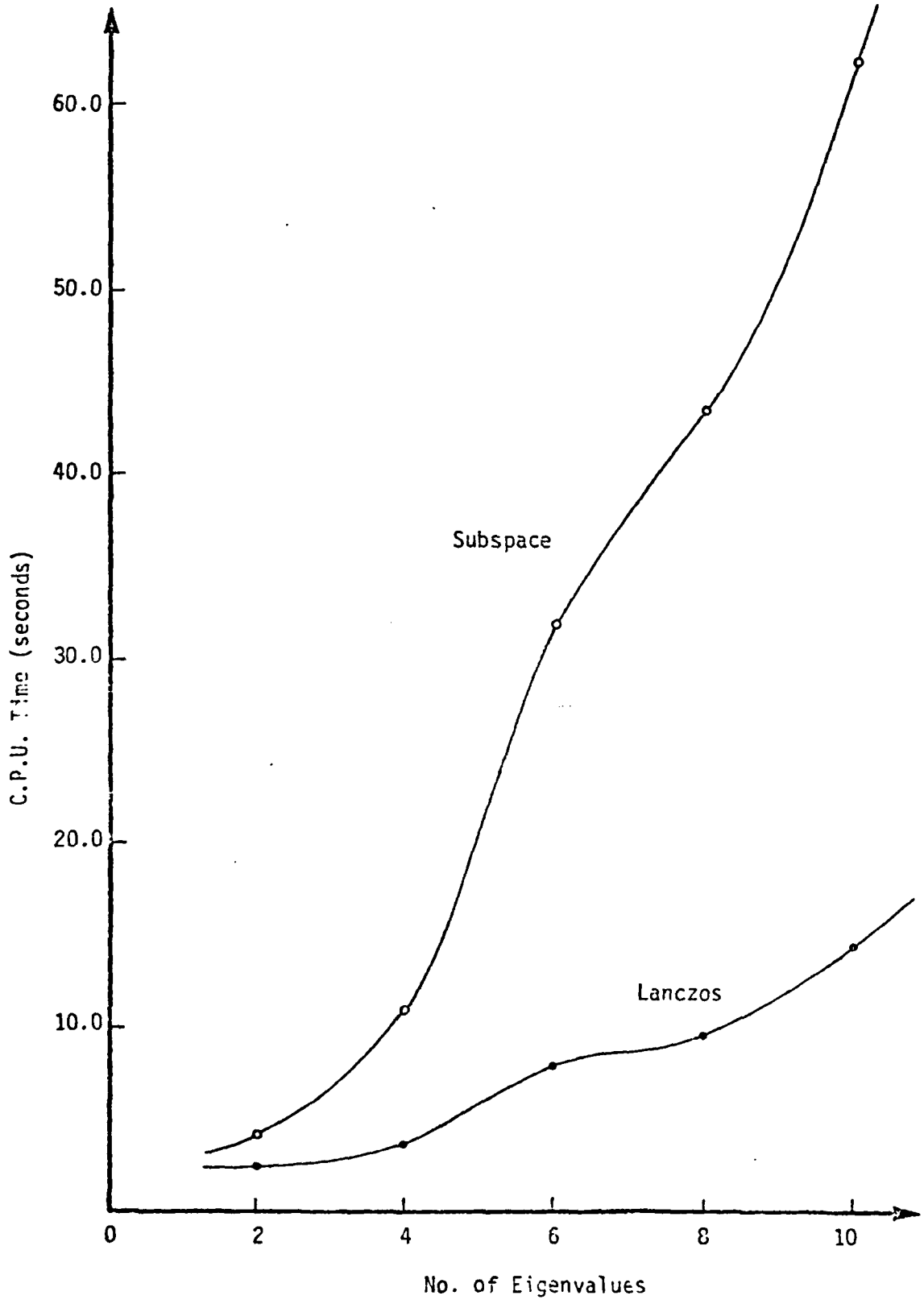
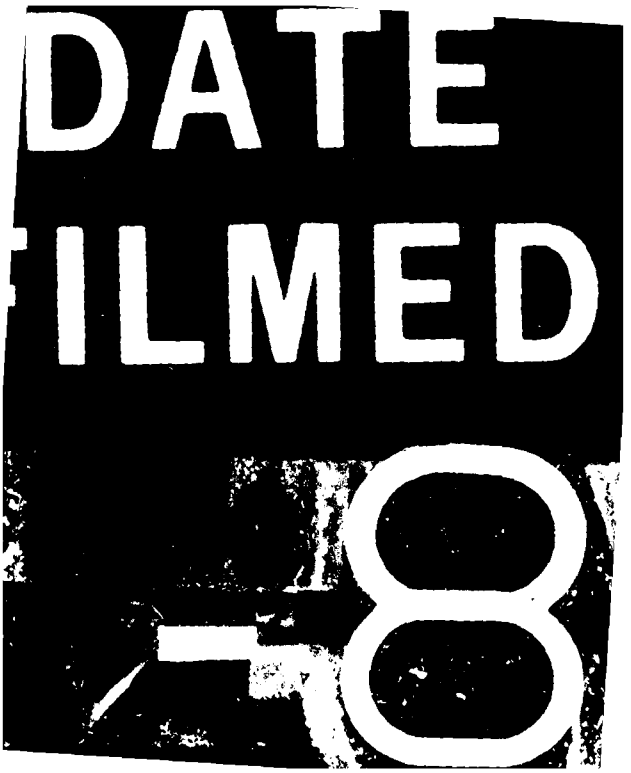


Figure 2. Comparison of Solution Times for Obtaining Increasing Number of Eigen pairs for Building Frame of 2.

This report was done with support from the Center for Pure and Applied Mathematics. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Center for Pure and Applied Mathematics or the Department of Mathematics.


DATE
ILMED
8